Teaching Case

Boosting Student Engagement and Analytics Skills Through Competition

Bryan Reinicke
breinicke@saunders.rit.edu

Quang Bui
qnbui@saunders.rit.edu

Victor Perotti
vperotti@saunders.rit.edu

Jing Tang
jtang@saunders.rit.edu

Rochester Institute of Technology
Rochester, NY

Hook

This case presents an analytics competition that has increased student engagement and understanding of the topic of analytics and interest in the College's degree programs in Business Analytics.

Abstract

Employers are increasingly looking for students with strong skills in business and data analytics. One of the challenges for faculty teaching analytics courses is that this requires pulling skills from multiple areas (statistics, programming, general business knowledge and systems thinking), and students frequently find the more technical aspects of analytics (statistics and programming) to be intimidating. This paper presents details of a competition that was designed to help boost student engagement and understanding of analytics. The competition helped the students build skills in this critical area and provided them with a concrete example to use while talking to prospective employers.

Keywords: Competition, Pedagogy, Analytics Skills, Student engagement

Boosting Student Engagement and Analytics Skills Through Competition

Bryan Reinicke, Quang Bui, Victor Perotti, Jing Tang

1. INTRODUCTION

Business and data analytics are among the top skills in demand by employers (Jafar et al., 2017). This has led to an increase in degree programs at both the undergraduate and master's level, and increased interest from students in the field. But these are also difficult skills for many students to learn.   The combination of technical skills (statistics and programming) and problem solving skills required to be successful are challenging to teach and are frequently introduced in multiple classes (Nestorov et al., 2019; Radovilsky & Hegde, 2022).

Competitions have been shown to be an effective way to increase students engagement with a topic (Lee et al., 2019; Pollacia et al., n.d.-a), and to improve their understanding (C J Chung, 2008). Giving the students a hands-on exercise using a realistic data set forces them to combine the various skills required and to understand how they are connected with one another.   The competition presented in this paper has been run for the last three years at the author's university, and has increased the number of participants each year.

By presenting information on the competition in our undergraduate and master's classes, we have also increased student awareness of the value of these skills and the classes offered in these areas.

2. LITERATURE REVIEW

While this is a teaching case, there is theory to support the approach presented here.

Business Analytics
Studies have shown that analytics is an in demand skill (Nestorov et al., 2019), in no small part because the skills are specialized.   This demand has dramatically increased student interest in the area, and has increased the number of degree programs in business analytics (Nestorov et al., 2019).

However, this increase has presented challenges for the faculty and universities building those programs (Nestorov et al., 2019). The skills required to be an effective data analyst are wide ranging (Collier & Powell, 2024) and are taught in a variety of different classes. Anytime it's necessary to link skills across courses, it presents a challenge for the faculty and the students. Hands on experience with problems like this can improve student learning (Mitri, 2023; Ye & Zhao, 2023), and competitions can provide that.

Competition
Prior research shows that competitions enhance student learning and engagement, particularly when they involve authentic, discipline-specific challenges (Pollacia et al., n.d.-b).

Engaging students with relevant business problems also improves their learning and understanding of how to apply their skills (Bashir et al., 2021; Stamper & Pardos, 2016), which is especially true when the skills are presented in multiple courses. Competitions are also a good example of active learning, which has been found to be a useful approach to teaching information systems concepts (Podeschi, 2016).

3. COMPETITION

The competition at our university was designed to serve a number of purposes.   First, it was intended to increase the learning opportunities for our students in the area of business analytics. Second, to increase their engagement with the business analytics community at large (more on this point later). Finally, it served as a way to increase student awareness of our degree programs in business analytics and the potential value of those degrees.

Initial Conceptual Design
The idea for the competition was first proposed in 2018, with the intention of creating a student Business Analytics competition within our University. The goal was to create a forum where students could apply classroom knowledge to solve real-world business analytics problems.

A second goal was to share an understanding of the practice Business Analytics, and how it differs from Data Science. Specifically, the competition emphasizes both the technical analysis including Machine Learning and the creation of

visualizations and communications to share business insights or recommend business action based on those insights. It was felt that this was particularly important, as it would serve to emphasize the need for various skill sets in business analytics professionals.

The idea was supported by the Dean Office and the faculty in the department responsible for business analytics. The competition was given financial support from the Deans office to pay for materials and prizes starting in 2021.

The First Competition
In the first year of the competition, it was open to all students at our university, both undergraduate and graduate. This allowed us to promote the competition across campus to all majors, including those in other non-business disciplines that could be interested in our graduate programs, as well as undergraduate minor in Business Analytics. It was determined that we should exclude PhD students from the competition, as they would likely have greater analytical skills than our other competitors and would not benefit as much from the experience. The students were required to form into teams of no more than five people.

The competition consisted of two rounds. In the first round, students have to solve a challenge and submit a business report of their solutions. The solutions are graded based on the technical sophistication of the solutions. The judges for the first round are professors with a background in business analytics. Additional details on the competition, including the flyer used to promote it, are presented in the Appendix.

In the second round, students present their solutions to another group of judges. The judges are business professionals and professors with business/technical backgrounds. Each student group was given a maximum of 25 minutes to present, with 10 minutes for Q&A.

All submissions were made through the in-house Learning Management System (LMS). Using the LMS provided all the judges with access to the files and an easy way of communicating with the students involved.

In the first year, the Competition attracted 9 submissions from student teams across four international campuses. There were over 100 students who attended our kick-off meeting events.

The Second Year Competition
In the second year, we made some adjustments to the competition based on the results and feedback from the first year. Feedback was gathered from discussions with the students who participated, the faculty involved and the industry experts who served as judges. The competition was again open to all undergraduate and master's students, but it was extended to teams outside of our own university as well. Students were again in teams, and the students were encouraged to form teams across schools to increase their exposure to different skills and a broader range of people. The information in the Appendices of this paper are from this year of the competition.

Similar to year one, the competition again consisted of two rounds. In the first round, students had to solve a challenge and submit a business report of their solutions. This year, another assessment criteria was added, in addition to faculty review. The solution also had to be submitted to Kaggle (www.kaggle.com) for a real-time evaluation of the solution. This addition provided two benefits to the competition. First, it engaged the students with an outside resource in the field of business analytics and presented them with a standardized score for their work. Second, it provided the faculty evaluating the competition with an outside score based on the students' work.

The solutions for the first round were graded based on the technical sophistication of the solutions (Kaggle score) + the quality of the business report as judged by the faculty running the competition. The judges for the first round were professors with a background in business analytics.

In the second round, students again presented their solutions to another group of judges. As in the first year of the competition, the judges for the second round were business professionals and professors with business/technical backgrounds and experience in business analytics.

As in the first year, all submissions were managed through the in-house Learning Management System (LMS) and in the Kaggle (www.kaggle.com) submission system.

Based on the feedback from the first year there was a workshop prior to the submission deadline for the first round, in which the students could interact with the first-round judges. The workshop allowed students to receive guidance and answers to their questions about what was expected. The workshop was run via Zoom to

allow students to participate regardless of their location.

In the second year, the Competition attracted 6 submissions from student teams, with one team consisting of non-University students.

4. LESSONS LEARNED

The team has learned a number of lessons over the last few years of running the competition, which will be valuable for anyone wanting to start a competition at their university.

First, be careful scheduling around campus wide events. One of the problems encountered in the first year of the competition was that we placed one of our deadlines too close to the campus wide career fair. This caused some problems as students were understandably more focused on getting a co-op or full-time job than on the competition.

Second, plan ahead. Ideally, running the competition early in the semester is best. That way, the students aren't yet bogged down by assignments and exams in their classes. The first year that we ran the competition, this wasn't possible as we were trying to get students in our intro business analytics course to participate. These students didn't have the skills at the beginning of the semester to complete the competition, so it had to be run later. When you're getting the competition off the ground, it can be incorporated into a business analytics course to ensure participation. Once the competition is established (after a few years of running it), moving it to earlier in the semester could increase student participation. Alternately, the first round of the competition can be run in the Fall semester, with the second round in the Spring semester.

Third, use outside experts as judges. Using judges from outside the university in the second round of the competition worked well for the students, and for the faculty. For some reason, the students seem to listen more closely to someone who is an industry professional, as opposed to a faculty member, even if they are saying the same thing. We recommend that you use the industry experts for the second round, so that the faculty members running the competition can give detailed feedback in the first round, which will improve the students' deliverables in the second round. This way, the students get feedback from your industry contacts on work that is more complete, and the industry contacts come away with a better overall impression of your students' abilities.

Fourth, have backups for the business judges. As we all know, sometimes things come up in the working world that can change schedules. As faculty, we have no control over the schedules for our industry contacts – so it's always a good idea to make sure that you have a couple of extra names of people who could judge the competition if something comes up.

Finally, having about 3 teams in the final round of the competition presenting to the external judges is ideal. Having too many teams in the final round will create problems with scheduling for the judges and the students. This also sets it up so that the students know from the beginning that only the best results will be able to go in front of outside judges and be eligible for the prizes. This helps to encourage them to put their best effort into the deliverables.

5. CONCLUSIONS

The competition worked as a way to increase student interest in learning business analytics skills, and to connect them with professionals in the area. The number of students participating has increased each year, and we have received positive feedback on the experience, both from them and the industry judges who have participated.

During the competition, the students were able to work directly with realistic data files and get feedback from industry professionals, which greatly increased their understanding of the process. By starting with a smaller dataset and receiving feedback on their work, they were able to develop both their skills and their confidence in their abilities before working on a larger one.

Running the competition in this way requires support from the local business community, the Dean's office and the faculty in the department. It's important to work with all of these stakeholders, as well as the students, to run a successful competition.

There is a good deal of work that goes into running a competition like this. However, it is worthwhile for everyone involved!

6. REFERENCES

Bashir, E., Nanath, K., & Hussain, F. (2021). Industry-academia outreach: A study of student perception. *International Journal of Technology Enhanced Learning*.

https://www.inderscienceonline.com/doi/10.1504/IJTEL.2021.118004

C J Chung. (2008). *CBL: Competition-Based Learning Learning through Competitions - Competition- Based Learning (CBL)*. https://doi.org/10.13140/RG.2.2.15673.90726

Collier, C., & Powell, A. (2024). Data Analyst Competencies: A Theory-Driven Investigation of Industry Requirements in the Field of Data Analytics. *Journal of Information Systems Education*, *35*(3), 325–376. https://doi.org/10.62273/SPYC4248

Jafar, M., Babb, J., & Abdullat, A. (2017). Emergence of Data Analytics in the Information Systems Curriculum. *Information Systems Education Journal*, *15*(5), 22–36.

Lee, P. T. Y., Lui, R. W. C., & Chau, M. (2019). *How Does Competition Help Future Learning in Serious Games? An Exploratory Study in Learning Search Engine Optimization*. *30*.

Mitri, M. (2023). *Using Python and AWS for NoSQL in a BI Course*.

Nestorov, S., Malliaris, M., & Jukic, N. (2019). Building a Business Data Analytics Graduate Certificate. *2019 International Conference on Computational Science and Computational*

Intelligence (CSCI), 859–864. https://doi.org/10.1109/CSCI49370.2019.00163

Podeschi, R. (2016). Building I.S. Professionals through a real-World Client Project ina Batabase Application Development Course. *Information Systems Education Journal*, *14*(6), 34–40.

Pollacia, L. F., Miller, P., Simpson, C., & McDaniel, N. (n.d.-a). *A Survey of Student Attitudes: Database Competition NCC 2003*.

Pollacia, L. F., Miller, P., Simpson, C., & McDaniel, N. (n.d.-b). *A Survey of Student Attitudes: Database Competition NCC 2003*.

Radovilsky, Z., & Hegde, V. (2022). *Contents and Skills of Data Mining Courses in Analytics Programs*.

Stamper, J., & Pardos, Z. A. (2016). The 2010 KDD Cup Competition Dataset: Engaging the machine learning community in predictive learning analytics. *Journal of Learning Analytics*, *3*(2), Article 2. https://doi.org/10.18608/jla.2016.32.16

Ye, C., & Zhao, L. (2023). *A Social Media Analytics Capstone Research Project with Community Engagement*.

Appendix A: Competition Announcements and Instructions

Annual Business Analytics Competition Spring 2022

In a competitive business world, companies are seeking students with both technical insight and business acumen to lead the charge to the future. To help students prepare for this competitive marketplace, the University College of Business presents the Business Analytics Competition, Spring 2022 that aims to bring together students from diverse backgrounds and skill levels to solve complex and meaningful business problems. This year, the competition is open to all students (not just University students) to foster a broader collaborative learning experience.

Scenario

Small Capital Bank (SCB) is a novel loan company headquartered in New York City. SCB loans are made available to almost anyone in small amounts, and can help those who otherwise would be unable to secure a loan from a more conventional bank. SCB has asked your team to advise them on three specific objectives:

1. Given the 2017-2018 loan portfolio, how well is SCB loan portfolio performing in regards to profitability? Which factors seem to be most associated with loan default and profitability?
2. Given the 2019 loan application data, create a decision rule for accepting/rejecting loan applications to maximize profit while avoiding systematic bias.
    a. The loan profitability calculations for SCB are described in the Appendix.
    b. When considering loan profitability for 2019 loan applications, students should think of innovative ways to estimate when a borrower will stop paying a loan (and subsequently becoming default). Some suggestion are:
        i. Using the 2017-2018 data to calculate the average time to default of a loan
        ii. Using the 2019 probability of default rate to calculate the profitability of a loan
3. Finally, SCB is concerned that its current loan acceptance rules have created unintended patterns, and therefore systematic bias. This may be in regard to borrowers due to their demographics (e.g., location, age, gender, marital status) or economic status (e.g., income class, home ownership). Thus, SCB is evaluating whether the current rules should be adjusted to avoid those unintended patterns while still maximizing market share and profitability.

Data

SCB have provided the following:

1. A dataset with information for all SCB's loans in 2017 and 2018 ("2022-dataset1.csv").
2. A dataset with information for SCB's loan applications in 2019 ("2022-dataset2.csv").
3. A data dictionary for all datasets ("2022-Data-Dictionary.xlsx")

Deliverables

Your team is asked to complete two deliverables:

1. Create a prediction model for loan defaults using the 2017-2018 dataset and predict the loan default probability for 2019 loan applicants (i.e., whether a given loan applicant will default or not). The prediction results will be uploaded to the Kaggle site for automatic evaluation.
2. Provide a summary report of no more than four pages (single spaced A4 size with 1" margin; excluding associated figures, visualizations, or tables) that communicates the answers to the two questions stated above by SCB. The report should contain the following elements:
    a. A cover page with names and school information of team members and contact information for the team leader. The cover page is not counted toward the page limit.
    b. An executive summary of key findings and recommendations
    c. A data preparation section that outlines the handling of data as well as providing details of any additional data used

     d. A data analysis section that reports the techniques and methodologies used by the team
     e. A results section that explains the findings and provides interpretations and recommendations for SCB's Board of Directors.
     f. An appendix section that contains any details the team wishes to clarify (e.g., visualizations, tables). The appendix section is excluded from the page limit.

All accompanying analysis files should be submitted together with the report (e.g., Tableau files, R/Python codes, Excel files, etc.).

Eligibility Criteria
1. Students can form teams of up to five students to participate in the competition.
2. The competition is open to both undergraduate and graduate students. Ph.D. students are not eligible for this competition.
3. Non-University students are welcome to participate. However, non-University students must create a University Guest account in University MyCourses site to submit their solution.
4. Participating students must currently enroll in an academic program at a university/college.

Procedures to Participate
1. Interested students should gain access the datasets and challenge in the competition website ( )
2. Students who are interested in submitting a solution to the competition need to request access to the competition course shell "Business Analytics Competition @ University" ( ). To gain access to the course shell, please send an email request to Prof. One ( ) or Prof. Two ( ).
     a. Non-University students who want to submit a solution need to create a guest account to University MyCourses system. Please send a request to Prof. One ( )
3. Students should form their own teams. Each team should have no more than five members.
4. There will be a workshop on March 23 @ 10am EDT, 2022 in which students will have the opportunity to ask questions and discuss their preliminary findings with some mentors.
5. Participating teams can submit their prediction model to Kaggle and summary reports to the MyCourses course shell. The deadline for submission is April 6, 2022 at 11:59pm EDT. Teams can submit as many times as they wish, but only the last submission will be used for evaluation. Late submissions will not be considered.
     a. Link to Kaggle submission site: https://www.kaggle.com/t/7e965e6b69c74c3f835836f7c3471e83

6. Names, school information, and contacts of team members should be included in the cover page of the submission. The main contact student for the team should be noted in the list. If not, the submitting student will be assumed as the main contact student.
7. The discussion board in MyCourses course shell will serve as a forum for questions and answers that students may have during the course of the competition. Students are encouraged to check the discussion board and post their questions there.
8. The evaluation of students' submissions will be done by April 12, 2022. The top four to six teams will be invited to make their presentations on April 22 @ 10am EDT, 2022. Each team will be given 15-20 minutes to present their findings to a panel of industry practitioners. The panel will select the top three teams as the prize winners for this competition.

Evaluation Criteria
In the first round (Kaggle model + report), each submission will be judged by a panel of instructors. The panel will assess the quality of each submission on the following dimensions:

1. **Technical accuracy and sophistication** - Does the technical work build confidence in the insights and recommendations?
2. **Business logic and implied reasoning** - Do the findings demonstrate an understanding of how the data contributes to the business decisions?
3. **Report coherency, narrative flow, and professionalism** - Does the summary report clearly communicate both the technical and business thinking to persuade an audience of BAC executives?

In the second round, the top four to six teams are invited to refine their understanding of SCB's problems and will make presentations on April 22, 2022. Each team will be given 15-20 minutes to present their findings to a panel of industry practitioners. The panel will use the following judging criteria:
1. Business logic/reasoning
2. Integration of oral and visual presentation elements
3. Quality of presentation/communication
4. Professionalism and persuasiveness
5. Responsiveness

Appendix: SCB Loan Profitability Calculations
SCB uses the following formula to measure loan profitability:

Loan Profitability = (Σ Loan Collect − Σ Loan Loss)/Loan Amount

For example, assume a loan of $100,000 with 24 months term at 10% interest rate. Using a loan calculator (https://calculator.me/loan/) will show that:
- The total amount will be collected from the loan (principal and interest) is $110,747.82. Thus, if the loan is successfully collected, the loan profitability will be:

  Loan Profitability = 110,747.82/100,000 = 1. 11

- If the loan borrower stops paying after the 22nd payment, and assuming that the lender will occur a $500 collection fee, the loan profitability will be:

  Loan Profitability = (amount received after 22nd payment) − (remaining principal + collection fee)/100,000
  =(110,747.82 − 9,114.95) − (9,114.95 + 500)/100,000 = 0. 92

In the context of SCB company:
Loan Profitability =(loan_total_payment + loan_late_fee_to_date − loan_remaining_principal −collection_fee)/loan_amnt

In general, a loan profitability greater than 1 is preferable, while loans with profitability less than 1 indicates potential performance issues.

APPENDIX B: Screen shots of Analytics Files

delinquent_accounts	borrower_state	annual_income	marital_status	default_12months	collection_fee
0	CA	42000	Single	0	0
0	VA	79077	Single	0	0
0	TX	107000	Married	0	0
0	NJ	155000	Single	0	0
0	TX	120000	Single	0	0
0	IL	32000	Single	0	0
0	NY	85000	Single	0	0
0	FL	90000	Single	0	0
0	FL	216000	Single	0	0
0	KY	10000	Married	0	0
0	IN	120000	Single	0	0
0	GA	88000	Single	0	0
0	NJ	58000	Single	0	0
0	MI	43000	Single	0	0
0	AZ	70000	Single	0	0
0	CA	13000	Single	0	0
0	VA	55000	Single	0	0
0	SC	52000	Single	0	0
0	OR	55000	Single	0	0
0	NY	62500	Single	0	0
0	TX	90000	Single	0	0
0	CA	139000	Single	0	0
0	WY	75000	Single	0	0
0	CA	47000	Single	0	0
0	VA	79000	Single	0	0
0	PA	89000	Single	0	0
0	WA	50000	Single	0	0
0	CT	80000	Married	0	0
0	TX	109000	Single	0	441.3618
0	FL	110000	Single	0	0
0	GA	65000	Single	0	0
0	WI	200000	Single	0	0
0	MA	90000	Single	0	0

Dataset 1: Approximately 212 MB CSV File

loanID	delinquen	borrower_	annual_in	marital_st	default_12	collection	delinquen	delinquen	loan_desc	employme
1	0	OK	50000	Single	1	0	1	0		8 years
2	0	FL	196000	Single	0	0	0	0		2 years
3	0	NH	44000	Married	0	0	1	0		< 1 year
4	0	AL	65000	Single	0	0	1	0		10+ years
5	0	WA	52000	Single	0	0	0	0		9 years
6	0	FL	52000	Single	0	0	0	0		10+ years
7	0	FL	19000	Married	0	0	0	0		
8	0	CA	36500	Married	0	3	3	0		5 years
9	0	OR	50000	Single	0	0	0	0		10+ years
10	0	GA	80000	Single	0	0	0	0		4 years
11	0	PA	46000	Single	0	0	0	0		1 year
12	0	WI	59940	Single	0	0	1	0		10+ years
13	0	GA	85000	Single	0	0	1	0		10+ years
14	0	NY	150000	Single	0	0	0	0		10+ years
15	0	WI	50000	Single	0	0	0	0		5 years
16	0	OR	72672	Single	0	1	0	0		10+ years
17	0	GA	55000	Single	0	0	0	0		10+ years
18	0	GA	82000	Single	0	0	0	0		10+ years
19	0	MD	160000	Single	0	0	0	0		7 years
20	0	NY	45000	Single	0	0	0	0		4 years
21	0	DC	41000	Single	0	0	0	0		4 years
22	0	CA	50000	Single	0	0	0	0		4 years
23	0	WA	39520	Single	0	0	1	0		< 1 year
24	0	TX	40000	Single	0	0	0	0		< 1 year
25	0	NY	55000	Single	0	0	0	0		< 1 year
26	0	PA	53700	Single	0	0	0	0		10+ years
27	0	NY	339000	Single	0	0	0	0		10+ years
28	0	LA	72000	Single	0	0	1	0		3 years
29	0	OR	40000	Married	0	0	0	0		6 years
30	0	MN	68000	Single	0	0	0	0		5 years
31	0	WA	100000	Single	0	0	0	0		10+ years
32	0	CA	120000	Single	0	0	1	0		10+ years
33	0	CA	225000	Single	0	0	0	0		4 years
34	0	NC	148000	Single	0	0	0	0		6 years

Dataset 2: Approximately 94 MB CSV File

APPENDIX C: Data Dictionary

Variable	Description
delinquent_accounts	The number of currently delinquent accounts of the borrower
borrower_state	The state of residence of the borrower
annual_income	The annual income of the borrower
default_12months	Number of defaults in the last 12 months
collection_fee	Collection fee for defaults
collections_12months	Number of collections in 12 months excluding medical collections
delinquent_2years	The number of delinquency in the borrower's credit file for the past 2 years
delinquent_amount	The amount owned by the borrower in all delinquent accounts
loan_description	Loan description
employment_length	Borrower's employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
employment_title	The job title of the borrower
fico_range_high	The upper boundary range the borrower's FICO
fico_range_low	The lower boundary range the borrower's FICO
loan_amount	The total amount for the loan
home_ownership	The home ownership status of the borrower
inquiries_finance	Number of personal finance inquiries
inquiries_credit_6months	The number of inquiries in past 6 months (excluding auto and mortgage inquiries)
loan_monthly_payment	The monthly payment for the loan
loan_interest_rate	Interest rate on the loan
loan_issue_date	The date when the loan was funded
loan_status	Current status of the loan
fi_mortgage_account	Number of mortgage accounts.
fi_late_account	Number of accounts with late payments
fi_active_debit_account	Number of currently active debit accounts
fi_total_debit_account	Number of total debit accounts
fi_active_credit_account	Number of currently active credit accounts
fi_total_credit_account	Number of total credit accounts
fi_active_all_account	The number of all active accounts in the borrower's credit file (e.g., debit, credit, store cards)
loan_remaining_principal	Remaining outstanding principal for the loan
derogatory_record	Number of derogatory public records
bankruptcy_record	Number of public record bankruptcies
loan_category	A category for the loan
loan_recovery	The amount of recovery from default
tax_lien_record	Number of tax liens

loan_term	The number of payments on the loan. Values are in months and can be either 36 or 60.
loan_title	The loan title
collection_total_amount	Total collection amounts ever owed by the borrower
fi_total_all_account	The total number of all accounts currently in the borrower's credit file (e.g., debit, credit, store cards)
loan_total_payment	Payments received to date for the loan
loan_interest_to_date	Interest received to date for the loan
loan_late_fee_to_date	Late fees received to date for the loan
loan_principal_to_date	Principal received to date for the loan
annual_income_verified	Whether the annual income was verified
borrower_zipcode	The first 3 numbers of the zip code of the residence of the borrower
marital_status	Marital status of the borrower
gender	Gender of the borrower (0=non-binary, 1=female, 2=male, 3=wish not to disclose)
age	Age of the borrower

Data Dictionary for the competition CSV Files

APPENDIX D: Scoring Rubric for First Round Deliverables

Criterion	Unsatisfactory 1	Satisfactory 3	Exemplary 5
Technical accuracy	Multiple inaccuracies were identified in the quantitative analysis of the data.	Some minor inaccuracies were identified in the presentation of findings.	Analytical findings accurately represent the underlying data set(s); no inaccuracies were identified.
Sophistication of Quantitative Analyses	The selection of analytical methods raises significant questions.	The quantitative analyses of the data reflect generally appropriate analytical methods; some methodological choices warrant additional justification.	The quantitative analyses of the data reflect appropriate and interesting application of analytical methods; team's analysis is innovative and creative.
Business logic/reasoning	Significant gaps in the business logic or reasoning of the submission are discernible.	The business logic of the report is generally coherent; some logical gaps or unstated assumptions are reflected.	The submission reflects a readily discernible business logic and reasoning about the data.
Integration of written summary(-ies) and quantitative analyses	The quantitative analyses are not well-integrated with the narrative report.	Some question regarding the integration of the written summary and the quantitative analyses were identified.	Written portions of the report are very well-integrated with the quantitative analyses presented.
Report coherency, narrative flow	Significant weaknesses in writing quality are identified; multiple typographical and grammatical errors are present; organization of the submission is difficult to follow; evidence of plagiarism exists.	Writing quality is generally strong; few typographical and grammatical errors are present; organization of submission is generally followable, but some re-organization may be needed.	The report is well-written and free of typographical and grammatical errors; all portions of the submission are well-organized and coherently presented.
Professionalism and Persuasiveness	Professional appearance of the submission - both written and quantitative portions - is relatively low; recommendations/findings are not well oriented to the intended audience; persuasiveness of findings is limited.	Submissions are generally well-polished; some revision of aesthetic appearance or presentation would be beneficial; recommendations are relatively persuasive.	Submission reflects professional polish (e.g., pages numbered, all graphics appropriately labeled); recommendations are relatable to the intended audience(s) and deemed insightful and persuasive.